Machine Learning Algorithms in Suicide Prevention:
Clinician Interpretations as Barriers to Implementation

Lily A. Brown, PhD[a,*]; Kathy Benhamou, BA[a,b]; Alexis M. May, PhD[c]; Wenting Mu, PhD[a]; and Richard Berk, PhD[d]

ABSTRACT

Objective: Machine learning algorithms in electronic medical records can classify patients by suicide risk, but no research has explored clinicians' perceptions of suicide risk flags generated by these algorithms, which may affect algorithm implementation. The objective of this study was to evaluate clinician perceptions of suicide risk flags.

Methods: Participants (n = 139; 68 with complete data) were mental health clinicians recruited to complete online surveys from October 2018 to April 2019.

Results: Most participants preferred to know which features resulted in a patient receiving a suicide flag (94.12%) and reported that knowing those features would influence their treatment (88.24%). Clinicians were more likely to report that some algorithm features (increased thoughts of suicide) would alter their clinical decisions more than others (age, physical health conditions; $\chi^2 = 270.84$, $P < .001$). Clinicians were more likely to report that they would create a safety/crisis response plan in response to a suicide risk flag compared to other interventions ($\chi^2 = 227.02$, $P < .001$), and 21% reported that they would complete a no-suicide contract following a suicide risk flag.

Conclusions: Clinicians overwhelmingly reported that suicide risk flags in electronic medical records would alter their clinical decision making. However, clinicians' likelihood of acting in response to a suicide risk flag was tied to which features were highlighted rather than the presence of the risk flag alone. Thus, the utility of a suicide risk algorithm will be reduced if clinical features underlying the algorithm are hidden from clinicians or if clinicians do not view the clinical features as intuitively meaningful predictors of suicide risk.

J Clin Psychiatry 2020;81(3):19m12970

To cite: Brown LA, Benhamou K, May AM, et al. Machine learning algorithms in suicide prevention: clinician interpretations as barriers to implementation. *J Clin Psychiatry*. 2020;81(3):19m12970.

To share: https://doi.org/10.4088/JCP.19m12970

[a]Department of Psychiatry, University of Pennsylvania, Philadelphia, Pennsylvania

[b]Department of Psychological Sciences, Case Western Reserve University, Cleveland, Ohio

[c]Department of Psychology, Wesleyan University, Middletown, Connecticut

[d]Department of Criminology and Department of Statistics, University of Pennsylvania, Philadelphia, Pennsylvania

**Corresponding author:* Lily A. Brown, PhD, 3535 Market St, Ste 600 N, Philadelphia, PA 19104 (lilybr@upenn.edu).

Suicide rates have increased over the past two decades.[1] A number of brief suicide prevention strategies are associated with reduced suicidal behavior, including safety planning,[2] Crisis Response Planning,[3] and the Coping Long-Term with Active Suicide Program (CLASP).[4] However, implementation of these strategies requires evidence-based methods for determining which patients are at the highest suicide risk. Risk assessment must incorporate data from a variety of sources, including objective and subjective indicators, as well as allow for complex interactions in risk factors. Electronic medical records (EMRs) offer an ideal platform for enhancing suicide risk detection because they offer longitudinal objective and subjective data.

Machine learning algorithms in EMRs can optimize prediction of suicide risk by combining longitudinal demographic, mental, and physical health data.[5] These algorithms may outperform traditional assessment to predict future suicidal behavior.[6] The Veterans Administration (VA) developed the Recovery Engagement and Coordination for Health—Veteran Enhanced Treatment (REACH VET) machine learning algorithm to identify veterans at risk for suicide using 381 features.[7] Promising preliminary findings from the first year of REACH VET implementation demonstrated reduced all-cause mortality.[7] Similarly, the Army Study to Assess Risk and Resilience in Servicemembers (STARRS) project used 421 features to classify active duty service members according to whether they died by suicide.[8] The Army STARRS and REACH VET projects found a variety of modifiable features (eg, substance use disorder, weapons possession) and unmodifiable features (eg, male sex, younger age, marital status) associated with suicide death. Some features had an intuitive connection to suicide risk (eg, prior suicidal ideation or behaviors, posttraumatic stress disorder), whereas others were less intuitive (eg, other anxiety-related disorders, nonaffective psychosis).[8,9] Regardless of the accuracy of an EMR machine learning algorithm in predicting future suicidal behavior, the success of the algorithm is entirely dependent on the decisions that clinicians make in response to suicide risk flags. Therefore, it is important to understand how clinicians interpret these algorithms and how their interpretations influence clinical decisions.

Several critical features of machine learning algorithms affect their interpretations. First, algorithms are not models, which are intended to represent some feature of how the world works; algorithms merely provide a computation. Second, because machine learning algorithms are not models, they do not allow for causal inferences. Third, features of a machine learning algorithm that drive classification into a high-risk group (eg, diagnostic or demographic factors) may not be causally related to suicide risk. Therefore, the features of a machine learning algorithm with the highest relative influence on classification may not be appropriate targets for clinical intervention. If a suicide risk flag is presented to clinicians alongside key algorithm features, clinicians may mistake these features as causing increased risk and correspondingly

Clinical Points

- Machine learning algorithms are being incorporated into electronic medical records to predict suicide, but no research has examined clinicians' interpretations of algorithm recommendations.
- Clinicians overwhelmingly reported that suicide risk flags in electronic medical records would alter their clinical decision making.
- Clinicians strongly preferred to know the clinical features that lead to a suicide risk flag from an algorithm.

change their treatment plan to target these features instead of targeting actual mechanisms of suicide risk. To our knowledge, there are no studies on clinician perceptions of and reactions to machine learning algorithms for suicide prevention. This is an important gap in the literature, because clinicians' interpretation of a suicide risk algorithm will very likely affect the extent to which suicide risk flags influence their clinical decision making.[10] Similarly, there are no studies on whether clinicians value knowing which features of a suicide risk algorithm drive classification, which might affect perception of algorithm credibility.[11] Finally, there are no studies on the interventions that clinicians would select when presented with a suicide risk flag. This choice is important because some interventions are supported by the literature, whereas others are contraindicated though commonly used.[12,13]

The first goal of this study was to understand whether clinicians preferred to know which algorithm features had the strongest influence on classification. We hypothesized that most clinicians would prefer knowing which features resulted in a suicide risk flag. The second goal was to understand whether clinicians reported that they would alter their clinical decision making because of the algorithm features. We hypothesized that participants would report that their clinical decision making would be altered based on knowing which features resulted in a suicide risk flag. The third goal was to understand whether certain features would be more likely to alter clinical decision making than others. We hypothesized that clinical features with high face validity (eg, suicide attempt history, suicidal ideation increase, family history of suicide) would be more likely to change reported clinical decisions relative to features with low face validity (eg, medical conditions, demographic features). The fourth goal was to understand whether there were differences in which suicide prevention interventions would be used most frequently in response to a suicide risk flag. We hypothesized that clinicians would report greater use of safety planning/crisis response planning following the provision of a suicide risk algorithm. The final goal was to understand clinicians' perceptions of the cost of false positives versus false negatives in a suicide risk algorithm. We hypothesized that most clinicians would report that a suicide risk algorithm should be designed to avoid a false negative at the expense of a false positive.

METHODS

The institutional review board of the University of Pennsylvania approved this study.

Recruitment

Participants were recruited via postings to Listservs (eg, the Association for Behavioral and Cognitive Therapies, local community mental health groups, the Zero Suicide Listserv) and social media platforms directed at mental health professionals from October 2018 to April 2019. Participants were entered a drawing for one $50 gift card.

Participants

Participants (n = 139) were recruited and completed the survey and informed consent form online. Eleven participants did not provide any additional information after providing informed consent and were dropped from all analyses. In addition, 60 participants provided demographic information but did not complete the surveys included in the current report. Therefore, 68 participants provided demographic information and surveys. There were no differences on demographic features of clinicians who completed the surveys versus those who did not, except for psychiatrists, who were significantly less likely to complete the survey (see Table 1). Participants were mostly female and white. There was diversity in representation of education and profession, though the majority had a master's degree and were licensed. The most common occupations were social workers and psychologists, though psychiatrists, nurses, and other allied professions were also represented. Specialty area was diverse, and less than half of participants indicated that suicide prevention was a specialty area.

Measures

Machine learning views. Participants were presented with two questions: (1) "In general, would you like to know which features [from a machine learning algorithm] led to a particular patient being flagged?" and (2) "In general, would knowing those features influence your treatment?"

Clinical features. Clinical feature selection questions were provided with the following prompt:

> Imagine that you were working with an electronic medical record system that could flag participants based on a machine learning algorithm. A machine learning algorithm takes large amounts of data across many people and uses statistics to make predictions about future behavior. This algorithm identifies individuals at risk for suicide, and provides a few clinical details describing why the individual received a suicide flag. For each indicator listed below, please indicate whether learning of this indicator through an electronic medical record system would change your clinical course of action. Please note that there are no right or wrong answers.

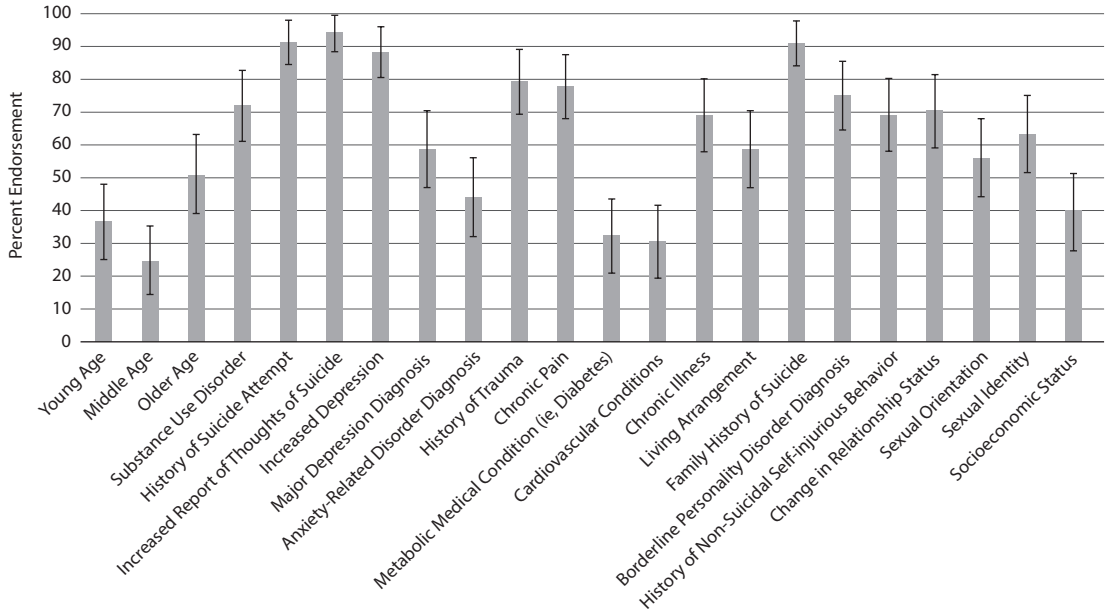
Participants were then presented with a list of indicators representing demographic, historical, and clinical features (see Figure 1) and were asked to indicate "Yes" or "No" for each.

Table 1. Participant Demographic and Professional Characteristics

Variable	Full Sample (n = 128), Mean (SD)	Completed Surveys (n = 68), Mean (SD)	Did Not Complete Surveys (n = 60), Mean (SD)	*P*
Age, y	37.86 (10.14)	37.87 (9.60)	37.85 (10.79)	.99
Years in practice	9.70 (8.85)	9.64 (8.03)	9.76 (9.76)	.94
Weekly hours of direct service	20.88 (13.19)	20.80 (12.30)	20.97 (14.26)	.94
	n (%)	n (%)	n (%)	
Gender				1.00
Female	109 (85.16)	58 (85.29)	51 (85.00)	
Male	17 (13.28)	9 (13.24)	8 (13.33)	
Transgender	0 (0.00)	0 (0.00)	0 (0.00)	
Prefer not to disclose	2 (1.56)	1 (1.47)	1 (1.67)	
Sex				.99
Female	111 (86.72)	59 (86.76)	52 (86.67)	
Male	17 (13.28)	9 (13.24)	8 (13.33)	
Race (multiple selections possible)				
White	112 (87.50)	61 (89.71)	51 (85.00)	.44
Black	5 (3.91)	1 (1.47)	4 (6.67)	.19
Latino	6 (4.69)	4 (5.89)	2 (3.33)	.68
Native American/Alaska native	4 (3.13)	1 (1.47)	3 (5.00)	.34
Asian	8 (6.25)	3 (4.41)	5 (8.33)	.47
Other	1 (0.78)	0 (0.00)	1 (1.67)	.47
Ethnicity				.86
Latino	8 (6.25)	4 (5.88)	4 (6.67)	
Non-Latino	120 (93.75)	64 (94.12)	56 (93.33)	
Degree				.84
High school diploma	1 (0.78)	0 (0.00)	1 (1.67)	
Associate degree	2 (1.56)	1 (1.47)	1 (1.67)	
Bachelor's	5 (3.91)	2 (2.94)	3 (5.00)	
Master's	85 (66.41)	45 (66.18)	40 (66.67)	
Doctorate	35 (27.34)	20 (29.41)	15 (25.00)	
Profession (multiple selections possible)				
Psychologist/psychology student	37 (28.91)	21 (30.88)	16 (26.67)	.46
Psychiatrist	4 (3.13)	0 (0.00)	4 (6.67)	.03
Nurse/nurse practitioner	8 (6.25)	5 (7.35)	3 (3.33)	.50
Social worker	41 (32.03)	25 (36.76)	16 (26.67)	.40
LPC	14 (10.94)	7 (10.29)	7 (11.67)	.80
Mental health counselor/behavior specialist	15 (11.72)	9 (13.24)	6 (10)	.57
Drug/alcohol counselor	5 (3.91)	3 (4.41)	2 (3.33)	.75
Marriage and family therapist	6 (4.69)	3 (4.41)	3 (5.00)	.88
Peer specialist/outreach worker	4 (3.13)	2 (2.94)	2 (3.33)	.90
Administrator/researcher	6 (4.69)	2 (2.94)	4 (6.67)	.13
Licensure status				.10
Licensed	96 (75.00)	55 (80.88)	41 (68.33)	
Unlicensed	32 (25.00)	13 (19.12)	19 (31.67)	
Specialty (multiple selections possible)				
Mood/personality disorders	69 (53.91)	41 (60.29)	28 (46.67)	.12
Anxiety disorders	73 (57.03)	42 (61.76)	31 (51.67)	.25
Suicide prevention	59 (46.09)	34 (50.00)	25 (41.67)	.35
Sleep disorders	7 (5.47)	4 (5.88)	3 (5.00)	.83
Substance use disorders	35 (27.34)	19 (27.94)	16 (26.67)	.87
Military/veterans	14 (10.94)	8 (11.76)	6 (10.00)	.75
Family therapy	12 (9.38)	5 (7.35)	7 (11.67)	.40
Couples therapy	10 (7.81)	5 (7.35)	5 (8.33)	.84
Child/adolescent	28 (21.88)	13 (19.12)	15 (25.00)	.42
Geriatric	3 (2.34)	3 (4.41)	0 (0.00)	.10
Women's health	13 (10.16)	7 (10.29)	6 (10.00)	.96
Men's health	4 (3.13)	3 (4.41)	1 (1.67)	.37
Culture	2 (1.56)	1 (1.47)	1 (1.67)	.93
LGBTQ health	12 (9.38)	8 (11.76)	4 (6.67)	.32
Pain/illness/medical	10 (7.81)	6 (8.82)	4 (6.67)	.65
Psychosis	5 (3.91)	4 (5.88)	1 (1.67)	.22
Neuropsychology	1 (0.78)	0 (0.00)	1 (1.67)	.29
Crisis intervention	3 (2.34)	0 (0.00)	3 (5.00)	.06
Eating disorders	2 (1.56)	0 (0.00)	2 (3.33)	.13
Community/public health	2 (1.56)	1 (1.47)	1 (1.67)	.93
Trauma/PTSD/domestic violence	7 (5.47)	4 (5.88)	3 (5.00)	.83

Abbreviations: LGBTQ = lesbian, gay, bisexual, transgender, and queer or questioning; LPC = licensed professional counselor; PTSD = posttraumatic stress disorder.

Figure 1. Clinical Features That Alter Behavior Following Suicide Risk Flag[a]



[a]Bars represent the frequency of endorsement (and 95% CI) of each clinical feature in response to the prompt, "Please indicate whether learning of this indicator through an electronic medical record system would change your clinical course of action."

Clinical interventions. Clinical intervention questions were provided with the following prompt:

> In general, in the event that you received a suicide flag and information explaining why that flag was triggered, would you employ any of these specific interventions (or target any of these specific factors) based on that information alone?

Participants were then presented with a list of potential interventions (see Figure 2) and were asked to select any that they would employ.

Perceptions of risk. Finally, participants were asked the following:

> In your opinion, which is worse: A false negative (i.e., a system that misclassifies or misses a true positive, in other words, a person who is at high risk for suicide is not flagged for suicide) or a false positive (i.e., a system that misclassifies a true negative, in other words, a person who is not at risk for suicide is flagged at risk for suicide?

They were then prompted to rate "how much worse" was their selection (a false positive or a false negative) on the following scale: 2 times, 3 times, 4 times, 5 times or more. This question was included to assess clinician preference for algorithm sensitivity versus specificity.

Data Analysis

A χ^2 test was conducted to evaluate the difference in perceptions of features that would alter clinical decision making in response to a suicide risk flag. The test was followed by an examination of the Pearson adjusted residuals, which follow a normal distribution.[14] For the clinical features analysis, residuals more extreme than ± 3.01 were indicative of more or less frequent endorsement, following Bonferroni correction for 19 items.[14–16] These analyses were repeated to evaluate the interventions that clinicians reported that they would use after a suicide risk flag, with a residual cutoff of ± 3.06 for a Bonferroni correction for 22 items.

RESULTS

Perceptions Toward Machine Learning

Most participants (94.12%) reported that they preferred to know which features resulted in a patient receiving a suicide flag, and the majority (88.24%) reported that knowing those features would influence their treatment.

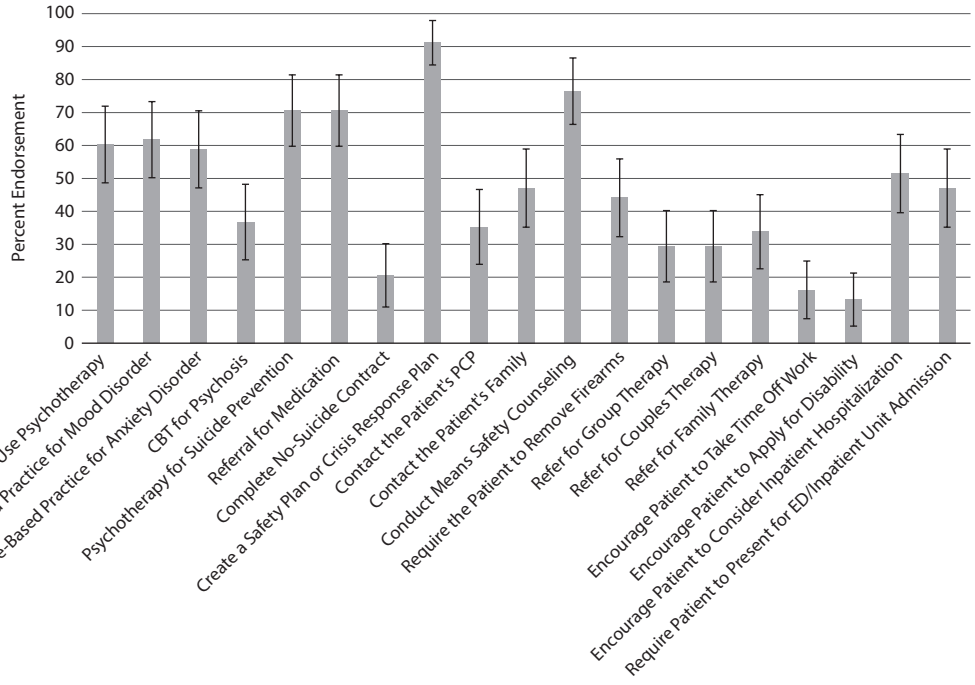
Algorithm Features That Alter Clinician Behavior

An omnibus χ^2 test of the likelihood of endorsing different features was significant ($\chi^2 = 270.84$, $P < .001$), such that clinicians were more likely to report that some features reported by the algorithm would alter their clinical decision making more than others (see Figure 1). Increased thoughts of suicide was selected most frequently (94.12%), and middle age (25.00%), cardiovascular conditions (30.88%), and metabolic conditions (32.35%) were selected least frequently.

Selected Interventions

An omnibus χ^2 test of the likelihood of endorsing different interventions following a suicide risk flag was significant ($\chi^2 = 227.02$, $P < .001$; see Figure 2). Of the 19 listed interventions, participants endorsed "creating a safety plan/crisis response plan" the most frequently (91.17%), followed by "conduct means safety counseling" (76.47%). The least frequently selected intervention was "encourage

Figure 2. Interventions Selected Following Suicide Risk Flag[a]



[a]Bars represent the frequency of endorsement (and 95% CI) of each clinical feature in response to the prompt, "In the event that you received a suicide flag and information explaining why that flag was triggered, would you employ any of these specific interventions (or target any of these specific factors) based on that information alone?"
Abbreviations: CBT = cognitive-behavioral therapy, ED = emergency department, PCP = primary care physician.

patient to apply for disability" (13.23%), followed by "encourage patient to take time off work" (16.18%). "Complete no-suicide contract" was selected by 20.59% of the sample.

Perceptions of Risk

Most participants (86.76%) indicated that a false negative was worse than a false positive. Of those who indicated that a false negative was worse, about half of those (55.93%) reported that a false negative was 5 times worse or more, with the remainder reporting that it was 4 times worse (10.17%), 3 times worse (18.64%), and 2 times worse (15.25%). Of participants who reported that a false positive was worse (13.24%), the majority reported that it was 2 times worse (66.67%), and the remainder indicated 5 times worse or more (33.33%).

DISCUSSION

Clinicians overwhelmingly reported that suicide risk flags in electronic medical records would alter their clinical decision making. This important finding highlights the influence of a suicide risk flag and may justify implementation of suicide risk flags alongside appropriate training in modern suicide prevention strategies. Most clinicians also reported that they preferred to know which clinical features resulted in the provision of a suicide risk flag. However, clinicians reported that they were more likely to change treatment plans in the face of certain algorithm features (eg, increased suicidal ideation) and not others (eg, medical conditions). This pattern was concerning, as the features that may be critical to a machine learning algorithm classification are quite likely distinct from clinical indicators that are obviously and functionally linked to suicide risk. Furthermore, in machine learning algorithms, features may interact in ways that complicate direct interpretation. Thus, clinicians' likelihood of acting in response to a suicide risk flag was tied to which features were highlighted rather than simply to the presence of the risk flag itself. These findings suggest that the utility of a suicide risk algorithm will be reduced if clinical features underlying the algorithm are hidden from clinicians or if clinicians do not view the clinical features as intuitively meaningful predictors of suicide risk.

Consistent with study hypotheses, increased suicidal ideation was endorsed as most important to clinicians in their decision making. Clinicians who are trained to conduct suicide risk assessments are often instructed to identify suicidal ideation, suicide plans, suicide intent, and history of suicidal behaviors. Therefore, it is not surprising that a report of increased suicidal ideation may be a well-rehearsed indicator of elevated risk. However, in recent research on machine learning algorithms for suicide death classification, some features with the greatest relative importance were not intuitively related to suicide risk (eg, verbal violence, nonaffective psychosis[8]). Further, other common features identified by algorithms were unmodifiable (eg, male sex,

young age, criminal offenses[8,9]), which may be less likely to change clinician behavior. Algorithms that identify non-intuitive or non-modifiable features as important may be less effective at changing clinician behavior, a clear cause for concern. In addition, algorithms weigh the importance of multiple, interacting factors in predicting risk such that an independent factor may be unimportant unless combined with a multitude of other factors, both face-valid and not.

These findings indicate that the implementation of a suicide risk flag algorithm requires careful consideration of the information that is revealed to clinicians alongside the flag. On the one hand, implementation of a "black box" algorithm without information about classification features may cause clinicians to ignore the flag altogether because of their preference for this information. On the other hand, transparency may lead to confusion in interpreting information generated by the algorithm. To optimize a suicide risk algorithm, clinicians need to understand that the features that drive classification may not be appropriate targets for suicide prevention. For instance, historical and demographic features may drive classification of a given participant into a high-risk category, but on determination of such classification, suicide prevention strategies must target empirically based mechanisms of suicide risk, such as social connection and hopelessness,[17,18] and not historical, diagnostic, or demographic features. Clinicians must understand the important distinction between the features that drive classification into a high-risk group versus the features that may reduce risk. Future research should explore whether providing clinicians with information about the features that drive risk classification improves clinical outcomes for patients, especially in light of challenges in interpretation of algorithm features in isolation.

For machine learning algorithms to alter treatment planning, algorithms must be implemented alongside ongoing training and consultation in their utilization and interpretation.[11] One key focus of this education should be on the enhanced accuracy of a machine learning algorithm versus traditional decision making in predicting future suicidal behavior. A second key focus of this training should be on appropriate interpretation of risk flags, particularly when these flags run counter to predetermined notions of the most substantial predictors of risk. According to the results of this study, clinicians will quite likely make decisions that are consistent with algorithm prediction only in cases in which the features driving the classification have high face validity. Their doing so defeats the purpose of the algorithm altogether.

Also consistent with study hypotheses, clinicians reported that their most likely interventions following a suicide risk flag were safety/crisis response planning or lethal means counseling. These findings were encouraging, as both interventions are associated with significant reductions in suicide risk.[2,3,19–21] However, one-fifth of clinicians reported that they would use a no-suicide contract following the provision of a suicide risk flag. This finding was alarming because, at worst, no-suicide contracts are associated with increased risk for self-harm[12] and, at best, these contracts have no evidence for effectiveness.[22] These findings provide further support that clinicians should be offered ongoing training and consultation in modern suicide interventions concurrent with the implementation of suicide risk algorithms.[23] Without training in evidence-based practices for suicide prevention, clinicians may elect to use ineffective or iatrogenic strategies to manage suicide risk following the receipt of a suicide risk flag.

Most clinicians reported that machine learning algorithms for suicide prevention should be programmed to avoid false negatives at the expense of false positives. Indeed, about half of clinicians who endorsed this belief reported that a false negative was 5 times riskier relative to a false positive. Clinicians may not trust black box recommendations in general, thus explaining their desire to understand the features that drive the algorithm recommendation. However, clinicians also asserted that if a black box recommendation is imposed on them, they want the recommendation to be extremely sensitive. The downside of programming hypersensitive algorithms is that clinicians may begin to ignore the flags altogether if they are overused. Additionally, if clinicians alter their behavior in response to highly sensitive suicide risk flags, the health care system may become strained by increased encounters focused on suicide prevention. Fortunately, at least 3 suicide prevention strategies are relatively brief, namely crisis response planning,[3] safety planning,[2] and CLASP.[4] Increasing adoption of these brief interventions may paradoxically decrease cost to the health care system in the long run by reducing psychiatric hospitalization or placement in resource-intensive psychotherapies for suicidal ideation. However, leadership should anticipate that while suicide risk flags will save the lives of their patients who may have been overlooked by traditional risk assessments, expenses on the health care system will most likely increase in the short run.

There are several limitations of the current study. First, the study relied on self-report data, which may differ from observational data. Second, clinicians with an interest in machine learning may have been more likely to participate in the study. However, not all study participants reported that they valued the input of a machine learning algorithm for suicide classification. Third, most participants were female and white, and results should be replicated in a more diverse sample. Fourth, to reduce participant burden, the researchers made a priori selections of a limited set of clinical features and interventions of potential interest. Their doing so may have biased results compared to purely qualitative open-ended data collection, which should be the focus of a subsequent study. Finally, data on the number of people who received recruitment materials for the study were not collected, and therefore we cannot make a determination about the reach of these recruitment methods.

In summary, these findings offer several considerations for the implementation of suicide risk flags in EMRs. Clinicians universally reported that a machine learning

algorithm would alter their decision making. However, clinicians also reported that they wanted to know the features that resulted in a patient classification as high-risk for suicide and that only 1 clinical feature (increased suicidal ideation) would alter their behavior. For machine learning algorithms to be useful, clinicians must receive ongoing training and consultation in their interpretation. While clinicians reported that they would use two evidence-based practices for suicide prevention most frequently, one-fifth of participants reported that they would use a potentially iatrogenic suicide prevention strategy following a suicide risk flag. Finally, clinicians overwhelmingly preferred sensitivity over specificity of a machine learning algorithm, which may burden the health care system and lead to desensitization to suicide risk flags. While machine learning algorithms offer a promising approach for improving suicide risk detection, implementation of these flags will require close consultation with clinicians. These important findings highlight the need for real-world studies on the influence of suicide risk flags in clinical practice.

Submitted: June 24, 2019; accepted October 30, 2019.

Published online: April 21, 2020.

Potential conflicts of interest: The authors report no financial or other relationship relevant to the subject of this article.

Funding/support: None.

REFERENCES

1. Stone DM, Simon TR, Fowler KA, et al. Vital Signs: trends in state suicide rates—United States, 1999–2016 and circumstances contributing to suicide—27 states, 2015. *MMWR Morb Mortal Wkly Rep*. 2018;67(22):617–624.
2. Stanley B, Brown GK, Brenner LA, et al. Comparison of the safety planning intervention with follow-up vs usual care of suicidal patients treated in the emergency department. *JAMA Psychiatry*. 2018;75(9):894–900.
3. Bryan CJ, Mintz J, Clemans TA, et al. Effect of crisis response planning vs contracts for safety on suicide risk in US Army soldiers: a randomized clinical trial. *J Affect Disord*. 2017;212:64–72.
4. Miller IW, Camargo CA Jr, Arias SA, et al; ED-SAFE Investigators. Suicide prevention in an emergency department population: the ED-SAFE Study. *JAMA Psychiatry*. 2017;74(6):563–570.
5. Linthicum KP, Schafer KM, Ribeiro JD. Machine learning in suicide science: applications and ethics. *Behav Sci Law*. 2019;37(3):214–222.
6. Dawes RM, Faust D, Meehl PE. Statistical prediction versus clinical prediction: improving what works. In: Keren G, Lewis C, eds. *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*. Hillsdale, NJ: Lawrence Erlbaum; 1993:351–367.
7. Matarazzo B. REACH VET: Recovery Eengagement and cCoordination for Hhealth—Veterans Enhanced Treatment: predictive analytics for suicide prevention. US Department of Veterans Affairs; 2018. https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/3527-notes.pdf. Accessibility verified March 19, 2020.
8. Kessler RC, Warner CH, Ivany C, et al; Army STARRS Collaborators. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study To Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*. 2015;72(1):49–57.
9. McCarthy JF, Bossarte RM, Katz IR, et al. Predictive modeling and concentration of the risk of suicide: implications for preventive interventions in the US Department of Veterans Affairs. *Am J Public Health*. 2015;105(9):1935–1942.
10. Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process*. 1991;50(2):179–211.
11. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. *JAMA*. 2018;319(1):19–20.
12. Drew BL. Self-harm behavior and no-suicide contracting in psychiatric inpatient settings. *Arch Psychiatr Nurs*. 2001;15(3):99–106.
13. Edwards SJ, Sachmann MD. No-suicide contracts, no-suicide agreements, and no-suicide assurances: a study of their nature, utilization, perceived effectiveness, and potential to cause harm. *Crisis*. 2010;31(6):290–302.
14. Agresti A. *Categorical Data Analysis*. 2nd Edition. New York, NY: Wiley; 2002.
15. Sharpe D. Your chi-square test is statistically significant: now what? *Practical Assessment, Research, and Evaluation*. 2015;20:8.
16. Chen S-Y, Feng Z, Yi X. A general introduction to adjustment for multiple comparisons. *J Thorac Dis*. 2017;9(6):1725–1729.
17. Joiner TE. *Why People Die By Suicide*. Cambridge, MA: Harvard University Press; 2005.
18. Klonsky ED, May AM. The Three-Step Theory (3ST): a new theory of suicide rooted in the "ideation-to-action" framework. *Int J Cogn Ther*. 2015;8(2):114–129.
19. Gunnell D, Fernando R, Hewagama M, et al. The impact of pesticide regulations on suicide in Sri Lanka. *Int J Epidemiol*. 2007;36(6):1235–1242.
20. Kreitman N. The coal gas story: United Kingdom suicide rates, 1960–71. *Br J Prev Soc Med*. 1976;30(2):86–93.
21. Mann JJ, Apter A, Bertolote J, et al. Suicide prevention strategies: a systematic review. *JAMA*. 2005;294(16):2064–2074.
22. Rudd MD, Mandrusiak M, Joiner TE Jr. The case against no-suicide contracts: the commitment to treatment statement as a practice alternative. *J Clin Psychol*. 2006;62(2):243–251.
23. York JA, Lamis DA, Pope CA, et al. Veteran-specific suicide prevention. *Psychiatr Q*. 2013;84(2):219–238.

Editor's Note: We encourage authors to submit papers for consideration as a part of our Focus on Suicide section. Please contact Philippe Courtet, MD, PhD, at pcourtet@psychiatrist.com.